Effects of Topology and Excluded Volume on Protein Denatured State Conformational Properties†

Christopher R. Smith, Natasa Mateljevic, and Bruce E. Bowler*

Department of Chemistry and Biochemistry, University of Denver, 2190 East Iliff Avenue, Denver, Colorado 80208

Received April 4, 2002; Revised Manuscript Received May 31, 2002

ABSTRACT: The conformational constraints on protein denatured states are of prime importance in modulating early events in protein folding. Although structural studies have demonstrated residual structure in protein denatured states, much remains poorly understood with regard to the conformational properties of this state. Here, we investigate topological effects on loop formation probabilities in denatured iso-1-cytochrome *c* by comparing histidine−heme binding affinities for histidines on the N- versus the C-terminal side of the heme. For histidines N-terminal to the heme (preceding cysteine 14), the polypeptide emerges from the edge of the heme and must simply fold over to bind to the heme. For histidines C-terminal to the heme (following histidine 18), the polypeptide emerges from the back side of the heme and must wrap around the heme for the histidine to bind to the heme. Thus, the steric constraints on this wrap-around topology are expected to be much more demanding than for the heme-edge topology of the N-terminal histidines. Evaluation of loop formation probabilities in 3 M guanidine hydrochloride, conditions that fully denature the variants studied, demonstrates that N-terminal histidine−heme loop formation is 10−25-fold more favorable than C-terminal histidine−heme loop formation, for similar loop sizes. A two-dimensional square lattice model indicates that excluded volume is important in this topological preference. These data provide direct evidence that denatured state topology affects contact probability, and thus probable folding pathways, in a disordered protein.

The mechanism by which the conformational search required to fold a protein is limited is of both fundamental and practical significance (*1*−*5*). A number of proposals have been put forth concerning how the conformational search is limited so that a protein folds in a biologically reasonable time (*2*). The presence of well-defined folding intermediates is one means to limit the search (*3*). The intrinsic downward energy bias for folding, or folding funnel model, also provides for efficient protein folding (*4*). The conformational search is also limited if the denatured state retains ordered structure (*5*). Both thermodynamic (*6*) and structural (*7*, *8*) studies support the presence of residual structure in denatured proteins. NMR studies on the drkN SH3 domain (*9*, *10*) demonstrate that the residual denatured state structure is more pronounced under nondenaturing conditions. NMR studies, on both the drkN SH3 domain (*10*) and reduced lysozyme (*11*) in water, are consistent with long-range hydrophobic clusters stabilizing residual structure in the denatured states of proteins. In fact, the topological features of the dominant conformers of the drkN SH3 domain in water are similar to those of the native state structure (*10*, *12*). Even under strongly denaturing conditions, staphylococcal nuclease retains the essential features of its native state structural topology (*13*, *14*). Thus, denatured state topological properties appear to be of prime importance for efficient protein folding.

Recently, analysis of the native state topology of proteins has demonstrated a strong correlation between high content of local structural contacts (low contact order) and fast folding (*15*−*17*). Presumably, this correlation results from the intrinsic conformational constraints on the denatured state of a protein. Interestingly, two-dimensional (2D) square lattice models indicate that when a loop consistent with the formation of a turn of helix or a β-turn is imposed, the highest probability second loop leads to helix or sheet formation (*18*). Comparison of the lattice model data with the results of random flight theory demonstrates that the preference for secondary structure formation is an excluded volume effect. Thus, structural preferences result from intrinsic conformational constraints on a disordered polymer. Recently, we studied denatured state histidine−heme loop formation for a set of iso-1-cytochrome *c* variants (*19*), providing loop formation probabilities for loops of 9−83 amino acids. Our technique of evaluating loop formation probabilities in denatured proteins is a thermodynamic approach and thus provides an important complement to the structural studies on protein denatured states. A 37 amino acid loop was most probable, with an affinity ∼5-fold higher than for smaller loops. This result seemed at odds with a link between denatured state conformational properties and the observed correlation between folding rates and native state contact order (*19*). In the present work, we look explicitly at the effects of topology and excluded volume on loop formation probability in a denatured protein. The results underscore

† This work was supported by NIH Grant GM57635-02 (to B.E.B.). The NSF (Grant 9977691) is acknowledged for support of the Beckman CEQ2000 DNA analysis system. C.R.S. acknowledges support from the Partners in Scholarship program at the University of Denver.

* To whom correspondence should be addressed. Phone: 303-871-2985. Fax: 303-871-2254. E-mail: bbowler@du.edu.

10.1021/bi0259249 CCC: $22.00
Published on Web 07/12/2002

the importance of these factors in defining contact formation preferences in a disordered polypeptide.

MATERIALS AND METHODS

Preparation of Iso-1-cytochrome c Variants. The AcH11, AcH4, AcH(−2), and AcH(−2)S2T variants (in the terminology AcH*X*,[1] Ac indicates N-terminal acetylation in vivo, H indicates histidine and *X* the sequence position of the histidine, and S2T indicates that Ser 2 is converted to Thr) were produced in the pRS/C7.8 vector using the method of unique restriction site elimination (*20*). The AcH*X* variants are derived from the AcTM variant (vida infra) by adding a single histidine at position *X* in the sequence. The pRS/C7.8 vector was produced by removing the 990 base pair *Xma*I/*Hin*dIII fragment from the 2.5 kb *CYC1* gene fragment contained within the pRS425/CYC1 vector (*21*) and ligating this fragment between the *Xma*I and *Hin*dIII sites of the pRS425 yeast shuttle vector (*22*). The 990 bp *CYC1* fragment was gel purified and isolated by the LiCl method (*23*) prior to ligation into pRS425. The *Xma*I/*Hin*dIII fragment of the *CYC1* gene contains the necessary control elements for expression of this gene in *Saccharomyces cerevisiae* (*24*). The *CYC1* fragment carrying the AcTM variant [T(−5)S, K(−2)L, H26N, H33N, H39Q, C102S] of iso-1-cytochrome *c* was used to prepare the pRS/C7.8 vector, for the work described here. This variant has had all histidines except the native state heme ligand, His 18, removed and is N-terminally acetylated in vivo (*19*). The restriction sites *Sac*I and *Sac*II upstream from the *CYC1* gene in the multiple cloning sequence of the pRS425 DNA were eliminated and restored in alternation to provide selection for mutations in the *CYC1* gene. Initial mutations to the AcTM pRS/C7.8 DNA used the *Sac*I+II− oligonucleotide 5′-d(GCCACCGCGCTG-GAGCTCCAGC)-3′ (underlined base eliminates the *Sac*II restriction site) and *Sac*II digestion of mutant pRS/C7.8 DNA isolated from *Mut*S (BMH 71-18) *Escherichia coli* cells to select for mutations to the *CYC1* gene, as described previously (*19*). Second mutations [AcH(−2)S2T) variant] were selected using digestion by the *Sac*I enzyme in conjunction with the *Sac*I−II+ oligonucleotide 5′-d(GCCACCGCGGT-GCAGCTCCAGC)-3′ (underlined base eliminates the *Sac*I restriction site). Mutations were characterized by DNA sequencing using a Beckman CEQ2000 DNA analysis system. Mutant DNAs were transformed into the *S. cerevisiae* GM-3C-2 cell line (cytochrome *c* deficient; *25*), the transformants characterized, and the variant iso-1-cytochromes *c* isolated and purified from transformed yeast, as described previously (*21*).

Stability Measurements. GdnHCl denaturation monitored by circular dichroism was carried out, as described previously (*26*). All measurements were made at 25 °C in the presence of 20 mM Tris, pH 7.0, and 40 mM NaCl, as buffer. The ellipticity as a function of [GdnHCl] was fit to a two-state unfolding model by nonlinear least-squares methods, assuming a linear free energy relationship for the dependence of unfolding free energy, ΔG_u, on [GdnHCl] (*26*). Values for the stability in water, $\Delta G_u(H_2O)$, and the denaturant dependence of ΔG_u (*m*-value) were obtained from the fits to the data. The errors reported are the standard deviation of three separate experiments.

Denatured State Titrations. Denatured state titrations were carried out in 3 M GdnHCl, 5 mM Na_2HPO_4, and 15 mM NaCl by monitoring the absorbance at 398 nm as a function of pH, as previously described (*19*, *26*−*28*). Data were obtained at 3 and 7.5 μM protein concentration for all variants. For the AcH11 variant, data also were obtained at 15 μM protein concentration. At 15 μM protein concentration the transition was monitored at 380 nm. The data were fit to eq 1 to obtain the observed pK_a, pK_a(obs), and the number

$$A_{398} = \frac{A_{398,LS} + A_{398,HS} \times 10^{n(pK_a(obs)-pH)}}{1 + 10^{n(pK_a(obs)-pH)}} \quad (1)$$

of protons involved in the process, *n*. In eq 1, A_{398} is the absorbance at 398 nm at a given pH, $A_{398,LS}$ is the absorbance at 398 nm of the low-spin form of the heme (histidine bound in the sixth coordination site), and $A_{398,HS}$ is the absorbance at 398 nm of the high-spin form of the heme (H_2O bound in the sixth coordination site). The values reported are the average of three separate experiments at 3 μM protein concentration and two separate experiments at 7.5 and 15 μM protein concentration. The error given is the larger of the standard deviation of the separate experiments or the error in the parameter from the nonlinear least-squares fit of the data obtained from the output of Sigmaplot (v.7).

For cases where the histidine−heme ligation is concentration dependent, the intermolecular (K_{inter}) and intramolecular (K_{intra}) equilibrium constant expressions are given by eqs 2 and 3, respectively. At the midpoint of the titration,

$$K_{inter} = \frac{[\text{cyt } c \text{ Fe}-\text{His cyt } c][H^+]}{[\text{cyt } c \text{ Fe}][\text{cyt } c \text{ HisH}^+]} \quad (2)$$

$$K_{intra} = \frac{[\text{cyt } c \text{ Fe}-\text{His}][H^+]}{[\text{cyt } c \text{ Fe}]} \quad (3)$$

$K_{inter} \times [\text{cyt } c \text{ HisH}^+] = [H^+]$ and $K_{intra} = [H^+]$. Since [cyt c HisH$^+$] = [cyt c]$_{total}$/2 at the titration midpoint, the concentration dependence of the observed equilibrium constant derived from the titration midpoint is given by eq 4.

$$K_{obs} = K_{intra} + (K_{inter}/2)[\text{cyt } c]_{total} \quad (4)$$

2D Square Lattice Models. The square lattices used for the two histidine−heme binding topologies are as defined in the caption to Figure 3. Exhaustive enumerations were made more efficient by taking advantage of the pseudo-2-fold symmetry of the square lattices. The set of conformations for the previous chain length, $n - 1$, were used as a starting point to enumerate conformations for the next chain length, *n*, by adding one monomer to the end of the chains of the $n - 1$ set in all orientations yielding no overlap with

[1] Abbreviations: GdnHCl, guanidine hydrochloride; AcH*X*, an iso-1-cytochrome *c* variant acetylated at the N-terminus, which contains a single surface histidine at position *X*; AcTM, a variant of iso-1-cytochrome *c* with no surface histidines and N-terminal acetylation; pK_a(obs), observed pK_a for loss of histidine−heme ligation under denaturing conditions; C_m, midpoint of guanidine hydrochloride unfolding titration; $\Delta G_u(H_2O)$, free energy of unfolding extrapolated to zero denaturant concentration; *m*, slope of the dependence of free energy of unfolding on guanidine hydrochloride concentration; P_x, fraction of conformations that form a loop of size *x*.

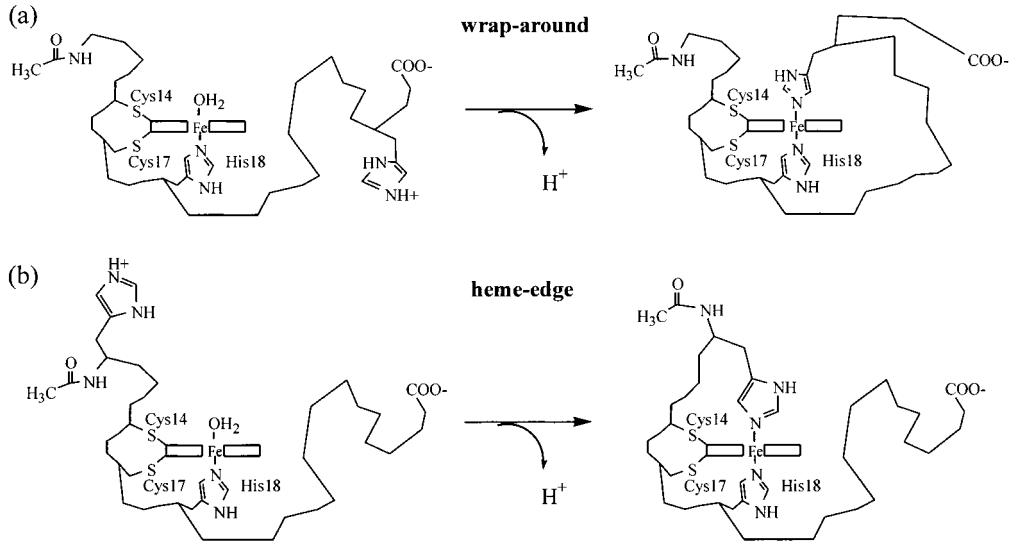

FIGURE 1: Schematic representation of the (a) wrap-around and (b) heme-edge topologies of histidine−heme binding in the denatured state of cytochrome *c*. In both (a) and (b) the low-pH form of the heme with water bound in the sixth coordination site is shown on the left. In the wrap-around binding mode histidines C-terminal to the site of heme attachment (Cys 14, Cys17, and His 18) to cytochrome *c* bind to the heme as the pH is raised. In the heme-edge topology, histidines N-terminal to the site of heme attachment bind to the heme as the pH is raised.

other chain monomers and consistent with the allowed 90° and 180° bond angles in the 2D square lattice.

RESULTS AND DISCUSSION

Loop Formation Topology and Variant Design. The asymmetry of the attachment of the heme to the polypeptide of cytochrome *c* provides a facile means to evaluate topological effects on denatured state loop formation. The heme is attached through thioether linkages to cysteines 14 and 17 at the edge of the heme plane (Figure 1). Under denaturing conditions, the native interaction between His 18 and the heme iron remains, whereas the Met 80 bond to the iron is lost and replaced by other ligands (*29*). Near neutral pH, the predominant heme ligands, under denaturing conditions, are the imidazole side chains of histidine (*19*, *26*−*31*). Inspection of Figure 1 shows that the steric constraints on loop formation will be very different for histidine−heme loop formation by histidines that are on the C-terminal versus the N-terminal side of the heme. Histidine side chains C-terminal to the heme attachment site will have a wrap-around topology (Figure 1a) for loop formation at the sixth coordination site of the heme (Met 80 binding site in the native state), since this portion of the polypeptide emerges from the back side of the heme. Histidine side chains N-terminal to the heme attachment site at Cys 14 will have a heme-edge topology (Figure 1b) for binding to the heme at the sixth coordination site, since this portion of the polypeptide emerges from the edge of the heme plane. Since heme is a large prosthetic group (∼8.5 Å diagonally across), it will significantly affect the sterics of the polypeptide chain. Thus, the heme of cytochrome *c* can serve as a model for the excluded volume effects of residual structure on the conformational properties of a denatured protein and the effects of chain topology with respect to that residual structure. Although the precise shapes are different, a heme is similar in dimension to ∼1.5 turns of α-helix or a 6-residue β-hairpin. Several β-hairpin structures are persistently populated in the ensemble of structures that represent the unfolded state of the drkN SH3 domain in water (*12*). Thus, the excluded volume effects of heme can provide an important benchmark for the effect of this type of persistent residual structure on the conformational properties of the adjacent polypeptide chain.

In previous work (*19*, *26*, *27*), we have studied denatured state histidine−heme binding for histidines C-terminal to the heme of iso-1-cytochrome *c*. Thus, to investigate topological effects on denatured state conformational properties, we have prepared variants of iso-1-cytochrome *c* with single histidines N-terminal to the heme. Since the heme is attached only 19 residues from the N-terminus, comparison will only be possible with the shorter loop sizes provided by the C-terminal AcH*X* iso-1-cytchome *c* variants studied previously (*19*). Thus, we prepared the iso-1-cytochrome *c* variants AcH11, AcH4, and AcH(−2), which introduce single histidines at positions 11, 4, and −2, respectively. In all cases, a surface-exposed lysine was replaced by histidine. The N-terminal acetylation in these variants is essential to prevent N-terminal amino group−heme ligation from interfering with evaluation of the histidine−heme equilibrium under denaturing conditions (*19*, *28*). These variants can form denatured state histidine−heme loops of 4, 11, and 16 amino acids, allowing comparison with the previously reported AcH27 and AcH33 variants which form denatured state histidine−heme loops of 10 and 16 amino acids, respectively (*19*).

Since the steric properties of proline, glycine, and β-substituted amino acids should significantly impact polypeptide conformational properties relative to alanine and other non-β-substituted amino acids (*32*), the percent content of these amino acids within the denatured state histidine−heme loops formed was considered in variant design. There are no prolines in iso-1-cytochrome *c* N-terminal to Cys 14. The percent glycine content of the AcH11, AcH4, and AcH(−2) histidine−heme loops is 0%, 9.1%, and 12.5%, respectively. Since it is impossible to create a 4-residue loop with ∼10% glycine content, no attempt was made to rectify this discrepancy. The percent β-substituted content of the AcH11,

Table 1: Thermodynamic Parameters from GdnHCl Denaturation of AcH*X* Iso-1-cytochromes *c* at pH 7 and 25 °C

variant	$\Delta G_u°(H_2O)$ (kJ/mol)	*m* [kJ/(mol·M)]	C_m (M)
AcTM[a]	16.4 ± 0.6	15.5 ± 0.8	1.06 ± 0.01
AcH11	17.0 ± 1.0	25.1 ± 1.4	0.68 ± 0.04
AcH4	8.5 ± 0.6	19.4 ± 1.2	0.44 ± 0.01
AcH(−2)	11.1 ± 0.5	20.8 ± 1.0	0.54 ± 0.01
AcH(−2)S2T	7.5 ± 0.7	23.2 ± 1.4	0.32 ± 0.01

[a] Data taken from ref *19*.

AcH4, and AcH(−2) loops is 25%, 18.2%, and 12.5%, respectively. To control for this difference, the serine at position 2 of AcH(−2) was mutated to threonine, producing the variant AcH(−2)S2T (18.8% β-substituted in the denatured state histidine−heme loop formed).

Guanidine Hydrochloride Denaturation Experiments. All variants were characterized by guanidine hydrochloride (GdnHCl) denaturation to verify that the proteins were unfolded under the conditions used to measure loop formation (3 M GdnHCl). The data are summarized in Table 1. The midpoint GdnHCl concentration, C_m, for unfolding of each protein is significantly below 1 M; thus all variants are denatured in 3 M GdnHCl, the conditions used for denatured state histidine−heme titrations. The denaturant unfolding parameters of the parent protein, AcTM (no histidines except His 18), are provided in Table 1 for comparison. All variants, except for AcH11, are strongly destabilized relative to the AcTM variant. In previous work, we have shown that heme ligands provided by side chains other than histidine do not compete significantly for heme binding at or below pH 7 [pK_a(obs) for AcTM is ∼7.5; see ref *19*]. Thus, under the pH 7 conditions used for GdnHCl denaturation, histidine will be the heme ligand at the sixth coordination site of the heme in the denatured state. The highly favorable histidine−heme loops formed for all but the AcH11 variant (vida infra) will strongly stabilize the denatured state in our GdnHCl denaturation experiments, possibly accounting for the significant loss in overall stability for the AcH4, AcH(−2), and AcH(−2)S2T variants relative to the AcTM variant. It is interesting to note that the *m*-values for these variants are typical of the wild-type protein [$m = 21.4 \pm 1.5$ kJ/(mol·M); see ref *26*] and the C-terminal AcH*X* variant with the shortest loop, AcH26 [$m = 19.4 \pm 0.8$ kJ/(mol·M); see ref *19*]. In our work on the C-terminal AcH*X* variants (*19*), we noted that the *m*-value decreased with increasing loop size. With these variants, the loops are small and leave the majority of the polypeptide, C-terminal to the heme, unconstrained.

Denatured State Loop Probability Measurements. The histidine−heme affinity under denaturing conditions is measured with a simple pH titration (Figure 2) in 3 M GdnHCl (*19*, *26*−*28*). The spin-state transition induced by switching from a weak field (H_2O) to a strong field (histidine) ligand (Figure 1) can be conveniently monitored at the heme Soret absorbance band (*19*, *26*−*30*). Titrations of histidine−heme ligation for AcH11 and AcH(−2) at 3 μM protein concentration (Figure 2) are ∼1 proton processes (Table 2), as expected (see Figure 1). To verify that histidine−heme binding is intramolecular, pK_a(obs) values were determined as a function of protein concentration for each variant. A decrease in pK_a(obs) with increasing protein concentration

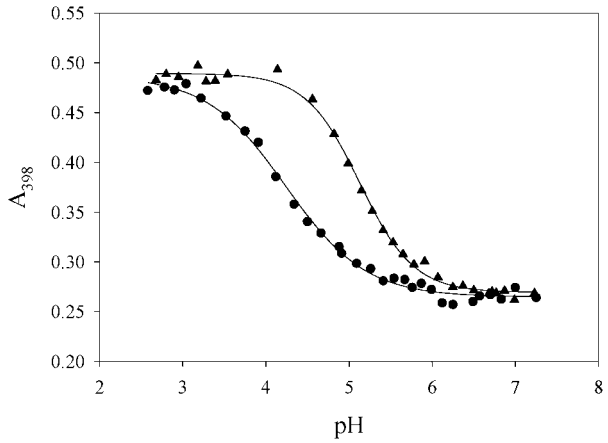

FIGURE 2: Titration of histidine−heme binding as a function of pH for N-terminal iso-1-cytochrome *c* variants. Histidine−heme binding is monitored at 398 nm in the Soret absorption band which is sensitive to the spin-state transition caused by switching from a strong field ligand (histidine) to a weak field ligand (H_2O) in the sixth coordination site of the heme. The solid triangles are data for the AcH11 variant at 3 μM protein concentration, and the solid circles are data for the AcH(−2) variant at 3 μM protein concentration. The solid curves are fits of the data to eq 1 in Materials and Methods. The parameters, from fitting the data, are collected in Table 2. All experiments were carried out at a [GdnHCl] of 3 M. Protein concentration was kept constant throughout the experiment. Data were obtained at 22 ± 1 °C.

Table 2: Equilibrium Parameters for Denatured State Loop Formation in 3 M GdnHCl[a]

variant	loop size[b]	[protein] (μM)	pK_a(obs)[c]	*n*[c]
AcH11	4	3	5.15 ± 0.03	1.39 ± 0.08
		7.5	5.02 ± 0.04	1.44 ± 0.11
		15	4.57 ± 0.06	0.98 ± 0.04
AcH4	11	3	4.22 ± 0.03	0.83 ± 0.07
		7.5	4.33 ± 0.05	0.98 ± 0.03
AcH(−2)	16	3	4.20 ± 0.04	0.95 ± 0.06
		7.5	4.15 ± 0.06	0.92 ± 0.28
AcH(−2)S2T	16	3	4.28 ± 0.09	0.92 ± 0.05
		7.5	4.29 ± 0.05	1.00 ± 0.04

[a] Titrations were carried out at 22 ± 1 °C. [b] Loop size is measured inclusively from Cys 14 to the histidine that binds to the heme in the denatured state. [c] The apparent pK_a, pK_a(obs), and the number of protons involved in the process, *n*, were obtained by fitting the pH versus absorbance data to eq 1 in Materials and Methods.

is expected if histidine−heme binding has an intermolecular component. This trend is observed only for the AcH11 variant (Table 2). A linear plot of K_a(obs) versus [AcH11] (eq 4, Materials and Methods) extrapolates to zero, within error, giving a value for the intramolecular histidine−heme equilibrium constant of $K_{intra} = (-0.1 \pm 5) \times 10^{-6}$. Thus, intramolecular histidine−heme binding is insignificant for AcH11. The intermolecular histidine−heme equilibrium, K_{inter}, derived from the slope of the K_a(obs) versus [AcH11] is 3.4 ± 0.9. Although the less sterically constrained heme-edge topology might have been expected to allow loop formation for the short 4-residue loop required by AcH11, such short loops are clearly unfavorable. The lack of the flexible amino acid, glycine (*32*, *33*), in the intramolecular loop required by AcH11, the small size of the loop, and the close approach of the two hemes in the intermolecular reaction probably all promote the intermolecular over the intramolecular reaction for AcH11.

The denatured state histidine−heme loops formed by the AcH4, AcH(−2), and AcH(−2)S2T variants are all highly favorable, giving denatured state histidine−heme pK_a(obs) values of 4.2−4.3 in 3 M GdnHCl (Table 2). For comparison, C-terminal AcH*X* variants with loop sizes of 9−22 gave pK_a(obs) values of 5.2−5.7 in 3 M GdnHCl. The 11-residue loop formed by the AcH4 variant is similar in size to the 10-residue loop formed by the AcH27 variant on the C-terminal side of the heme. The pK_a(obs) for the AcH27 variant is 5.68 ± 0.03 (3 M GdnHCl; see ref *19*), ∼1.4 p*K* units (25-fold) less favorable than the pK_a(obs) for the 11-residue loop formed by the AcH4 variant (Table 2). The 16-residue loop formed by the AcH(−2) and AcH(−2)S2T variants is the same size as the loop formed by the C-terminal AcH33 variant. The pK_a(obs) of 5.26 ± 0.04 (3 M GdnHCl; see ref *19*) for the AcH33 variant is ∼1 p*K* unit (10-fold) less favorable than the 16-residue loop formed by the AcH(−2) and AcH(−2)S2T variants (Table 2). Thus, for similar denatured state loop sizes, the heme-edge topology of N-terminal loop formation is considerably more favorable than the wrap-around topology of C-terminal loop formation.

Since the portion of the iso-1-cytochrome *c* polypeptide N-terminal to the heme is short (N-terminal amino acid is at position −5), it is not possible to obtain the broad range of loop sizes measured on the C-terminal side of the heme (*19*). Thus, the most favorable loop size is not known. For the C-terminal histidine variants, the AcH54 variant (loop size of 37) has the most favorable denatured state histidine−heme loop. Although calculations on poly(L-alanine) indicate that chain stiffness is important out to chain lengths of 40 (*34*), more recent calculations suggest a most probable loop size of ∼10 residues for loop formation in a disordered polypeptide (*35*). Given the different topologies for histidine−heme loop formation under denaturing conditions for N- versus C-terminal AcH*X* variants, it is conceivable that the most probable loop size would differ. Thus, comparison of the histidine−heme affinity of the N-terminal AcH*X* variants with the histidine−heme affinity of the C-terminal variant with the most favorable loop, AcH54, would provide a lower limit for the enhancement in loop formation probability for the heme-edge versus the wrap-around topology. The pK_a(obs) for the AcH54 variant is 4.81 ± 0.06 (3 M GdnHCl; see ref *19*), 0.5 p*K* unit (3-fold) less favorable than loop formation for either the AcH4 or AcH(−2)S2T variant. Thus, even at the lower limit, histidine−heme loop formation under denaturing conditions is significantly more favorable for the heme-edge versus the wrap-around topology.

In comparing denatured state histidine−heme affinities, it is also important to consider differences in the sequence composition of the loops formed, since this factor may influence loop formation probability. As noted above, the percent content of glycine in the histidine−heme loops formed by AcH4, AcH(−2), and AcH(−2)S2T is ∼10%. For the C-terminal AcH27, AcH33, and AcH54 variants, the glycine content is ∼20% in the histidine−heme loops formed. The greater flexibility of glycine (*32*, *33*) should enhance denatured state loop formation probability for the C-terminal versus the N-terminal variants. There are no prolines on the N-terminal side of the heme, whereas the proline content of the C-terminal denatured state loops formed by AcH33 and AcH27 is ∼10% and for AcH54 ∼5%. Insertion of individual prolines into a protein is expected to decrease the average dimension of a disordered polypeptide chain (*36*, *37*), again favoring denatured state loop formation for the C- versus the N-terminal AcH*X* variants. Like the AcH4 and AcH(−2)S2T variants, the percent content of β-substituted amino acids in the histidine−heme loops formed is ∼20% for the C-terminal AcH27, AcH33, and AcH54 variants. Overall, the sequence composition differences between the N- and C-terminal loops should favor loop formation from the C-terminal side of the heme, if anything, leading to an underestimate of the preference for heme-edge topology indicated by our pK_a(obs) measurements in 3 M GdnHCl.

The pK_a(obs) values for the AcH(−2) and AcH(−2)S2T variants provide an opportunity to experimentally evaluate the effect of β-substituted amino acids on the conformational properties of a disordered polypeptide. The β-substituted amino acids are expected to expand the dimension of a random coil (*32*), lowering loop formation probability. However, the effect is predicted to be modest (*37*). The pK_a(obs) data (Table 2) for AcH(−2) versus AcH(−2)S2T (12.5% versus 18.8% β-substituted, respectively) do show an increase in the pK_a(obs) of ∼0.1 p*K* unit consistent with the expected decrease in denatured state loop formation probability when an additional β-substituted amino acid is inserted in the 16-residue loop. However, the observed difference is not outside the error of our measurement. Thus, our experimental data indicate that the effect on the conformational properties of a disordered protein of β-substituted amino acids relative to non-β-substituted amino acids is of negligible importance.

Does the heme of cytochrome *c* perturb the properties of its denatured state, from a thermodynamic standpoint, compared to the denatured state of non-heme-containing proteins? Myers et al. (*38*) surveyed a broad range of proteins and examined two important thermodynamic parameters that describe the characteristics of the denatured state of a protein relative to the native state of a protein, denaturant *m* values and the heat capacity increment of unfolding, ΔC_p. Both of these parameters are expected to correlate with the change in solvent-exposed surface area upon unfolding, ΔASA. Good linear correlations were observed between ΔC_p and ΔASA and denaturant *m* values and ΔASA for the set of 45 proteins examined. In both cases, the data for cytochrome *c* fell right on the correlation line. Thus, thermodynamically, the denatured state of cytochrome *c* is representative of protein denatured states, and the perturbation due to heme appears to be negligible. Heme is sterically larger compared to many amino acids. Here we take advantage of this property, as described above, to model the effects of persistent residual structure on chain conformational properties. These effects are explored more fully with lattice models in the next section.

Loop Formation with a 2D Square Lattice Model. To gain further insight into topological effects on denatured state loop formation, we use a 2D square lattice model. This model has proven useful in determining loop formation probabilities in polymers and in understanding the role of excluded volume in modulating conformational properties (*18*). The 2D square lattice tends to accentuate excluded volume effects; however, qualitatively, the same conclusions are obtained with a 3D cubic lattice model (*18*, *39*). In our 2D square lattice model, three monomers are held fixed to

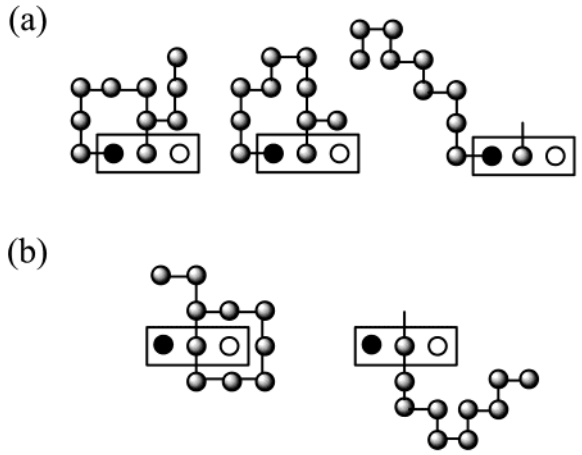

FIGURE 3: Representative 2D square lattice conformations for (a) the heme-edge topology for histidine−heme binding and for (b) the wrap-around topology for histidine−heme binding. The heme is represented by a black rectangle containing three monomers, represented as circles. The solid and open circles at opposite ends of the heme unit represent the inherent asymmetry of the heme (cysteine thioether side and heme propionate side, respectively). The line emerging from the central gray monomer represents the sixth coordination site of the heme where the loop is closed. Chain residues are shown as solid gray circles. The black lines are the bonds between the monomers. For the heme-edge topology, the chain emerges from the left side of the cysteine thioether monomer of the heme unit. The first chain monomer for the heme-edge topology is equivalent to Cys 14 in cytochrome *c*. For the wrap-around topology, the chain emerges from the central monomer of the heme unit on the side opposite from the sixth coordination site. The first chain monomer for the wrap-around topology is equivalent to His 18 in cytochrome *c*. Chains of nine monomers are shown for both the heme-edge and the wrap-around topologies. For the heme-edge topology, the diagram on the left shows a conformation with a 6-residue loop closed, the diagram in the middle shows a conformation with an 8-residue loop closed, and the diagram on the right shows a conformation with no loop closed. For the wrap-around topology, the diagram on the left shows a conformation with a 7-residue loop closed, and the diagram on the right shows a conformation with no loop closed. Note that only chain monomers, not heme monomers, are counted as part of the loop. The results of exhaustive enumeration of conformations on the heme-edge and wrap-around topology 2D square lattices are summarized in Table 3.

represent the heme (Figure 3). For the heme-edge topology (Figure 3a), the chain extends from one edge of the heme, and for the wrap-around topology (Figure 3b), the chain emerges from the side opposite the histidine binding site on the heme. We have carried out exhaustive enumeration of the possible conformations for chain lengths ranging from 4 to 10 monomers and tabulated the fraction of histidine−heme loop-forming conformations, P_x, where x is loop size (Table 3). Due to the nature of the 2D square lattice, the wrap-around topology can only form loops containing an odd number of chain monomers, and the heme-edge topology can only form loops containing an even number of chain monomers. Thus, loops that differ by one chain monomer, for the two topologies, will be compared.

There are obvious differences between the data for the heme-edge and the wrap-around topology. For the same chain length, the size of conformational space is significantly larger for the heme-edge versus the wrap-around topology (Table 3). That this is an excluded volume effect resulting from the topology of the chain with respect to the heme can be demonstrated in a straightforward manner by considering the

Table 3: Loop Formation Probabilities from 2D Lattice Models

Heme-Edge Topology				
chain length[a]	total conformations	P_4[b,c]	P_6[b,c]	P_8[b,c]
4	25	0.0400 ($l = 0$)		
5	67	0.0299 ($l = 1$)		
6	189	0.0265 ($l = 2$)	0.0212 ($l = 0$)	
7	509	0.0275 ($l = 3$)	0.0118 ($l = 1$)	
8	1399	0.0264 ($l = 4$)	0.0093 ($l = 2$)	0.0114 ($l = 0$)
9	3795	0.0271 ($l = 5$)	0.0095 ($l = 3$)	0.0071 ($l = 1$)

Wrap-Around Topology			
chain length[a]	total conformations	P_7[b,c]	P_9[b,c]
4	21		
5	55		
6	155		
7	417	0.00480 ($l = 0$)	
8	1163	0.00344 ($l = 1$)	
9	3093	0.00323 ($l = 2$)	0.00452 ($l = 0$)
10	8357	0.00335 ($l = 3$)	0.00287 ($l = 1$)

[a] Chain length is the number of monomers excluding the three representing the heme. [b] P_x is the fractional probability of forming a loop containing x monomers. Only chain residues, not heme residues, are counted as part of the loop. P_x is calculated by dividing the number of conformations that form a loop of size x by the total number of conformations possible for the chain length being enumerated. [c] The number, l, in parentheses indicates the number of monomers separating the monomer which forms the loop from the free end of the chain.

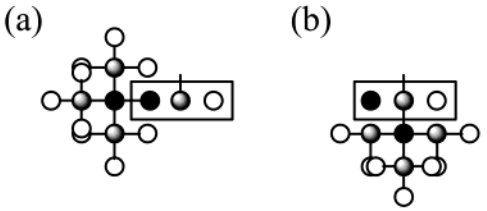

FIGURE 4: Relative excluded volume effects for the heme-edge (a) versus the wrap-around (b) topology in the 2D lattice model. The heme is represented as described in the caption to Figure 3. The first chain monomer is shown as a solid black circle. The possible conformations for the second monomer are shown as gray circles. The possible conformations for the third monomer are shown as open circles. The solid lines represent the bonds between the monomers.

conformational possibilities for the third monomer added to each chain (Figure 4). A maximum of three positions is possible for a monomer added to the end of a chain in a 2D square lattice. For the wrap-around topology, two of the three conformations for the second monomer only allow two conformations for the third monomer because of the presence of the heme (Figure 4b). In the case of the heme-edge topology, all three conformations for the second monomer permit three conformations for the third monomer (Figure 4a). This difference in the excluded volume effects for the wrap-around versus the heme-edge topology reduces the size of conformational space for the wrap-around topology relative to the heme-edge topology.

Inspection of Figure 4 shows that the disallowed conformations for the third monomer will tend to bias the conformational distribution for the wrap-around topology toward conformations on the back side of the heme. On the other hand, it is clear from the symmetry of the heme-edge topology that no such heme-sidedness bias will exist. Thus,

excluded volume plays a more prominent role in limiting conformational space for the wrap-around topology, decreasing the proportion of conformations on the histidine binding site side of the heme. From Table 3, it can be seen that the probability of forming loops with the histidine binding site of the heme is much lower for the wrap-around topology. Loops less than 7 monomers are not possible for the wrap-around topology, whereas loops as small as 4 monomers can form for the heme-edge topology. We observe end effects in the magnitude of P_x, as previously (*18*). It is easier to form a loop if the monomer closing the loop is the last residue in the chain (i.e., if the number of residues, l, separating the monomer that closes the loop from the end of the chain is zero; see Table 3). Most of the end effect is dissipated when the loop-forming monomer is one in from the end ($l = 1$), with smaller adjustments as the loop-forming residue moves further in from the end. For loops that are one residue in from the end of the chain ($l = 1$), the magnitudes of P_6 and P_8 for the heme-edge topology are 3.4 and 2.1 times larger, respectively, than P_7 for the wrap-around topology (see Table 3). Thus, the same qualitative difference is observed for the 2D square lattice model as observed experimentally, favoring the heme-edge topology over the wrap-around topology for loops of similar size. The quantitative differences are much larger in the pK_a(obs) measurements for loops of similar size. However, the lower limit for the increased favorability of the heme-edge topology obtained by comparing the pK_a(obs) values for the N-terminal variants with the pK_a(obs) value for the most probable loop size (37 amino acids, AcH54) with the C-terminal variants is similar to the prediction of the 2D square lattice model (both indicate that the heme-edge topology is favored over the wrap-around topology by ~3-fold).

In a 3D cubic lattice, the heme can be represented by a two-dimensional 3 by 3 square array of monomers with the chain emerging from the central monomer of the array (opposite to the histidine binding site) for the wrap-around topology. For the heme-edge topology, the chain would emerge from a central monomer on the edge of the 3 by 3 heme array. For a 3D cubic lattice, a maximum of five conformations is possible for a monomer added to the end of a chain. For the heme-edge topology, only two out of five conformational positions for the second chain monomer are in contact with the heme. These two conformational positions will only permit four out of five possible conformations for the third monomer. For the wrap-around topology, however, four of the five conformations for the second monomer place this monomer against the surface of the heme, limiting the third monomer to four out of five possible conformations, in all four cases. Thus, qualitatively, the excluded volume effects in three dimensions will be the same as in two dimensions, only the magnitude of the effect will be diminished, as noted previously (*18*, *39*).

The key insight from the 2D lattice model is that excluded volume plays a significant role in the differences between the heme-edge and the wrap-around topologies. The heme more effectively sterically blocks access to the histidine binding site in the wrap-around topology. Interestingly, the topology of the linkage to the heme also reduces the total number of conformations accessible for the wrap-around topology, simplifying the conformational search for an ordered structure, while at the same time skewing the distribution of possible conformations away from long-range (wrap-around) and toward local interactions.

Implications for Protein Folding. In previous work on the distance dependence of loop formation (*19*), we obtained a scaling factor of -4.2 ± 0.3 for loop formation in the C-terminal AcH*X* variants (wrap-around topology). This value is much higher than the freely jointed random coil value of -1.5 typically used to represent the conformational properties of denatured proteins (*40*). When excluded volume is accounted for, the scaling factor is expected to be -2.1 to -2.4 (*39*, *41*). Experimental measurements on small peptides are consistent with scaling factors of -1.5 (*42*−*44*). In these cases, it is unlikely that excluded volume effects due to residual structure are important. Combined data from a number of proteins for disulfide loop formation indicated a scaling factor of -2.4 (*45*), consistent with the predicted effect of excluded volume on random coil properties. The 2D square lattice models for the wrap-around versus the heme-edge topology suggest that excluded volume effects are more pronounced for the wrap-around topology of our C-terminal AcH*X* variants. Since excluded volume increases scaling factors, the chain topology relative to the heme may contribute to the large scaling factor observed for the C-terminal AcH*X* variants. Unfortunately, pK_a(obs) values are only available for two intramolecular loop sizes (11 and 16; see Table 2) for the N-terminal AcH*X* variants due to the short length of polypeptide N-terminal to the site of heme attachment. Thus, it is not possible to determine a scaling factor from the experimental data for the heme-edge topology.

Since the heme, as noted above, is large enough to represent a small element of persistent residual structure such as a β-hairpin, in a protein denatured state, our experimental and lattice model data suggest that the topology of the remaining disordered chain with respect to the element of residual structure can lead to significant biases in the conformational space of the remainder of the chain. The excluded volume effects of the residual structure can skew the conformational distribution, making scaling factors much steeper for some types of contacts for a given topology, as is observed for the wrap-around topology for histidine−heme binding in the denatured state of the C-terminal AcH*X* variants (scaling factor of -4.2 ± 0.3; see ref *19*). Topologically, the steric constraints of wrap-around histidine−heme binding are similar to those of a 5-residue β-hairpin structure (see Figure 5a). The excluded volume limits the choices for positioning the sixth monomer, favoring extension of the β-sheet and disfavoring longer range wrap-around contacts. On the other hand, the heme-edge topology is like a stretch of 4 monomers in β-structure with no turn. In this case, the limitations on the fifth and sixth monomers are less (Figure 5b), and the structural possibilities for the chain are greater and less skewed. Our data for the N- versus the C-terminal variants of iso-1-cytochrome *c* reflect these differences in conformational possibilities for polypeptides with topologically different residual structure and thus in a sense provide an experimental verification of the effects of imposed topological constraints on the conformational properties of polypeptide chains previously suggested by lattice models (*18*, *39*).

The skewing of the conformational distribution away from high contact order for loop formation for the wrap-around

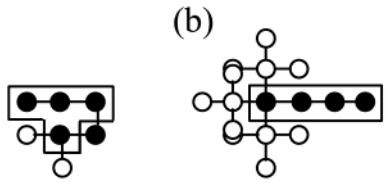

FIGURE 5: Two-dimensional lattice representations of the conformational possibilities for hexamer chains with residual structure. (a) A 5-residue β-hairpin residual structure fragment showing the conformational possibilities for the 6th residue. The residues corresponding to constrained residues in the wrap-around heme topology are boxed in. (b) A 4-residue segment of residual β-structure showing the conformational possibilities for the 5th and 6th residues. The residues corresponding to constrained residues in the heme-edge topology are boxed in. In (a) and (b), the residues that are part of the residual structure are shown as solid black circles, and the possible positions of the conformationally unconstrained residues are shown as open circles. The bonds between the monomers are shown as solid black lines.

topology provides a rationale in terms of denatured state conformational properties for the effects of native state contact order on folding efficiency. These topological constraints presumably depend on the exact nature and extent of the residual structure. However, recent structural studies on the denatured state of proteins under folding conditions indicate that the conformational biases caused by residual structure in a protein denatured state can be substantial (*9*−*14*). Our experimental loop formation measurements and lattice modeling results indicate that such conformational biases are attributable to the effects of excluded volume.

CONCLUSION

The data presented here provide a unique experimental demonstration of the importance of topology in modulating the probability of contact formation in a disordered polypeptide. The heme-edge topology of contact formation for denatured iso-1-cytochrome *c* is significantly favored over the wrap-around topology. Our data provide direct experimental validation that the probability, and thus the efficiency, of finding a native structure with a more local topology (low contact order) starting from a disordered state with residual structure will be higher than for a native structure requiring more long-range or wrap-around topologies. Many denatured proteins appear to retain substantial residual structure (*6*−*14*). Besides the obvious reduction in conformational space that results from residual structure (*4*), our results suggest that the precise topological constraints imposed by the residual structure, and the resulting excluded volume effects, will determine the degree of reduction in conformational space (Table 3) and will dictate what structural interactions are most likely to occur next in the process of folding a protein.

REFERENCES

1. Dobson, C. M. (1999) Protein misfolding, evolution and disease, *Trends Biochem. Sci. 24*, 329−332.
2. Levinthal, C. (1968) Are there pathways for protein folding? *J. Chim. Phys. 65*, 44−45.
3. Englander, S. W. (2000) Protein folding intermediates and pathways studied by hydrogen exchange, *Annu. Rev. Biophys. Biomol. Struct. 29*, 213−238.
4. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat. Struct. Biol. 4*, 10−19.
5. Dill, K. A., and Shortle, D. (1991) Denatured states of proteins, *Annu. Rev. Biochem. 60*, 795−825.
6. Shortle, D. (1995) Staphylococcal nuclease: A showcase of *m*-value effects, *Adv. Protein Chem. 46*, 217−247.
7. Shortle, D. (1996) Structural analysis of non-native states of proteins by NMR methods, *Curr. Opin. Struct. Biol. 6*, 24−30.
8. Dyson, H. J., and Wright, P. E. (1998) Equilibrium NMR studies of unfolded and partially folded proteins, *Nat. Struct. Biol. 5*, 499−503.
9. Zhang, O., and Forman-Kay, J. D. (1997) NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions, *Biochemistry 36*, 3959−3970.
10. Mok, Y.-K., Kay, C. M., Kay, L. E., and Forman-Kay, J. (1999) NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions, *J. Mol. Biol. 289*, 619−638.
11. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein, *Science 295*, 1719−1722.
12. Choy, W.-Y., and Forman-Kay, J. D. (2001) Calculation of ensembles of structures representing the unfolded state on an SH3 domain, *J. Mol. Biol. 308*, 1011−1032.
13. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures, *J. Mol. Biol. 268*, 170−184.
14. Shortle, D., and Ackerman, M. S. (2001) Persistence of native-like topology in a denatured protein in 8 M urea, *Science 293*, 487−489.
15. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins, *J. Mol. Biol. 277*, 985−994.
16. Mirny, L., and Shakhnovich, E. (2001) Protein folding theory: From lattice to all-atom models, *Annu. Rev. Biophys. Biomol. Struct. 30*, 361−396.
17. Fersht, A. R. (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism, *Proc. Natl. Acad. Sci. U.S.A. 97*, 1525−1529.
18. Chan, H. S., and Dill, K. A. (1989) Intrachain loops in polymers: Effects of excluded volume, *J. Chem. Phys. 90*, 492−509.
19. Hammack, B. N., Smith, C. R., and Bowler, B. E. (2001) Denatured state thermodynamics: Residual structure, chain stiffness and scaling factors, *J. Mol. Biol. 311*, 1091−1104.
20. Deng, W. P. D., and Nickoloff, J. A. (1992) Site-directed mutagenesis of virtually any plasmid by eliminating a unique site, *Anal. Biochem. 200*, 81−88.
21. Bowler, B. E., May, K., Zaragoza, T., York, P., Dong, A., and Caughey, W. S. (1993) Destabilizing effects of replacing a surface lysine of cytochrome *c* with aromatic amino acids: Implications for the denatured state, *Biochemistry 32*, 183−190.
22. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H., and Hieter, P. (1992) Multifunctional yeast high-copy-number shuttle vectors, *Gene 110*, 119−122.
23. Favre, D. (1992) Improved phenol-based method for isolation of DNA fragments from low-melting temperature agarose gels, *BioTechniques 13*, 22−26.
24. Guarente, L., and Mason, T. (1987) Regulatory proteins in yeast, *Annu. Rev. Genet. 21*, 425−452.
25. Faye, G., Leung, D. W., Tatchell, K., Hall, B. D., and Smith M. (1981) Deletion mapping of sequences essential for in vivo transcription of the iso-1-cytochrome *c* gene, *Proc. Natl. Acad. Sci. U.S.A. 78*, 2258−2262.
26. Godbole, S., Hammack, B., and Bowler, B. E. (2000) Measuring denatured state energetics: Deviations from random coil behavior and implications for the folding of iso-1-cytochrome *c*, *J. Mol. Biol. 296*, 217−228.
27. Godbole, S., and Bowler, B. E. (1997) A histidine variant of yeast iso-1-cytochrome *c* that strongly affects the energetics of the denatured state, *J. Mol. Biol. 268*, 816−821.
28. Hammack, B., Godbole, S., and Bowler, B. E. (1998) Cytochrome *c* folding traps are not due solely to histidine-heme ligation: Direct demonstration of a role for N-terminal amino group-heme ligation, *J. Mol. Biol. 275*, 719−724.
29. Babul, J., and Stellwagen E. (1971) The existence of heme-protein coordinate-covalent bonds in denaturant solvents, *Biopolymers 10*, 2359−2361.

30. Colón, W., Wakem, L. P., Sherman, F., and Roder, H. (1997) Identification of the predominant non-native histidine ligand in unfolded cytochrome *c*, *Biochemistry 36*, 12535−12541.
31. Muthukrishnan, K., and Nall, B. T. (1991) Effective concentrations of amino acid side chains in an unfolded protein, *Biochemistry 30*, 4706−4710.
32. Tanford, C. (1968) Protein denaturation, *Adv. Protein Chem. 23*, 121−282.
33. Miller, W. G., Brant, D. A., and Flory, P. J. (1967) Random coil configurations of polypeptide copolymers, *J. Mol. Biol. 23*, 67−80.
34. Conrad, J. C., and Flory, P. J. (1976) Moments and distribution functions for polypeptide chains, Poly-L-alanine, *Macromolecules 9*, 41−47.
35. Camacho, C. J., and Thirumalai, D. (1995) Theoretical predictions of folding pathways using the proximity rule, with application to bovine pancreatic trypsin inhibitor, *Proc. Natl. Acad. Sci. U.S.A. 92*, 1277−1281.
36. Mutter, M. (1977) Macrocyclization equilibria of polypeptides, *J. Am. Chem. Soc. 99*, 8307−8314.
37. Miller, W. G., and Goebel, C. V. (1968) Dimensions of protein random coils, *Biochemistry 7*, 3925−3935.
38. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant *m* values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding, *Protein Sci. 4*, 2138−2148.
39. Chan, H. S., and Dill, K. A. (1990) The effects of internal constraints on the configurations of chain molecules, *J. Chem. Phys. 92*, 3118−3135.
40. Betz, S. F. (1993) Disulfide bonds and the stability of globular proteins, *Protein Sci. 2*, 1551−1558.
41. Redner, S. (1980) Distribution functions in the interior of polymer chains, *J. Phys. A 13*, 3525−3541.
42. Eaton, W. A., Muñoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R., and Hofrichter, J. (2000) Fast kinetics and mechanisms in protein folding, *Annu. Rev. Biophys. Biomol. Struct. 29*, 327−359.
43. Lapidus, L. J., Eaton, W. A., and Hofrichter, J. (2000) Measuring the rate of intramolecular contact formation in polypeptides, *Proc. Natl. Acad. Sci. U.S.A. 97*, 7220−7225.
44. Bieri, O., Wirz, J., Hellrung, B., Schukowski, M., Drewello, M., and Kiefhaber, T. (1999) The speed limit for protein folding measured by triplet−triplet energy transfer, *Proc. Natl. Acad. Sci. U.S.A. 96*, 9597−9601.
45. Darby, N. J., and Crieghton, T. J. (1993) Dissecting the disulphide-coupled folding pathway of bovine pancreatic trypsin inhibitor: Forming the first disulphide bonds in analogues of the reduced protein, *J. Mol. Biol. 232*, 873−896.

BI0259249